401

6/21/2012

**Session 1**

Learning Objectives

After this session, the student will be able to:

**Chapter 1: • Describe the relationships between prediction, co-incidence (correlation), and causation.**

Prediction: Another term for a prediction is a hypothesis or hypotheses, which means – Statements suggesting possible relationships or associations among the phenomena being studied Pg. 5. Also, Hypothesis: A statement that names the variables that appear to be related and indicates the nature of that relationship Pg. 10.

Correlation: A demonstrated association or relationship between two variables, this could be either a positive relationship or inverse relationship.

Causation: A relationship where one variable brings about the other variable.

A prediction seeks to find a relationship between two variables or phenomenon. This relationship can be unrelated, correlated, or in a causal relationship.

In order for a relationship to be causal it must be correlative, but a correlating relationship does not need a causal relationship to be present 23.

Important Definitions:

Induction: A process of reasoning that goes from the specific to the general 11.

Deduction: A process of reasoning that goes from the general to the specific 12.

Positive Relationship: A relationship in which greater is associated with greater; less with less 19.

Inversely Related: A condition in which most cases cluster about the off diagonal, a high score on one variable is associated with a low score on the other 20.

Cause: When one phenomenon being studied brings about the other 23.

Temporal Sequence: When one phenomenon being studies occurs earlier in time than the other 23. (Thus, the former event is the causal ((Independent Variable)) factor in the experiment.

Dependent Variable: The variable that is being caused or explained 24.

Independent Variable: The variable that is doing the causing or explaining 24.

Unit of Analysis: What we actually measure or study to test our hypothesis.

**Chapter 2.Describe different forms of data and levels of measurement.**

Forms of Data:

* Qualitative or Categorical Data: Data that are assigned to categories that do not imply amounts, 35.
* Quantitative Data: Data that are assigned to categories that are involved with amounts, 35.

Levels of Measurement, 35:

* Nominal: Pertains to the act of naming, such as by attributes: characteristics that do not necessarily involve amounts, such as one’s gender, eye color, or religion.
  + Individual Nominal Data: Data that are presented in a list of each individual subject in the study and his or her assignment for one or more variables 36.
    - ie, Name Gender Religion

Bill Male Protestant

Jill Female Atheist

* + Grouped Nominal Data: Data that are presented as a category of the variable listed, and the subjects are not named but are counted in the category into which the each subject falls.
    - ie, Gender f=

Male 4

Female 6

n= 10

The above is a frequency distribution, 39.

* Ordinal: Involving a rank order or other ordering, 40.
  + Individual ordinal data: rank each individual subject from highest to lowest along the variable.
    - Name Height

Jack Tallest

Jill Second

James Shortest

* + Grouped Ordinal Data: Data that present subjects placed into ranked categories, ordered highest to lowest or inverse, 41.
    - Height f=

Very Tall 3

Tall 7

Short 2

Very Short 8

N= 20

* Interval: An interval scale in which the subject receives a numerical score rather than a ranking and where zero is an arbitrarily chosen point rather than the lack of what is being measured. Score can be above and below zero, 46.
  + Individual internal level data, or Raw score: A simple numerical score
    - Name Age
    - Bill 15
    - Jim 24
    - Will 30
  + Ungrouped frequency distribution: Scores listed in a sequence that includes every score that actually appears in our results, 49.
    - Score F=
    - 15 3
    - 14 0
    - 13 1
    - 12 1
    - N= 5
  + Grouped Interval data: Grouped data that are also at the interval level of measurement, 51. There are two criteria:
    - Class intervals must all be equal in size.
    - All class intervals must be closed-ended
  + 5
* Ratio: A level of measurement similar to interval level, but where zero is an absolute zero, meaning none of what is being measured. Scores can NOT be below zero.

Important Definitions:

Measurement: A very specific process, such as measuring length, but also other simpler actions such as assignment of a person to a particular category of a variable, 34.

Dichotomy: Two-category variables, such as Male or Female, 39.

n category: A term indicating more than two categories, 40.

Likert Scale: A scale whose categories are based on the level of agreement with a particular statement or issue, 43.

Absolute zero: A zero that means a complete lack of the variable being measured, 47.

Demographic Variables: Background information on the human subjects studied, 47.

Class Interval: An interval that indicates the space between two end points, 51.

**• Explain the importance of defining variables in statistics.**

Variables in statistics are the subjects in what we desire to know more about conduct experiments. In an experiment, when defining a variable it is best to start in a deductive fashion. Starting with a general definitions and working towards a precise working definition for a variable. Ideally, utilizing face validity will assist precise variables. Clearly communicating and thinking through the steps of an experiment will aid in the validity of variables.

Important definitions:

Demographic Data: Background information that fives the social characteristics, 65.

Working or Operational Definition: A definition of the way that someone or something will be measured to determine the subject’s score on a variable, 66.

Conceptual Definition: A general definition of a concept such as one would find in a textbook or dictionary, 67.

Items: The various components (abortion, family, values, etc.) used to generate a scale or index, 71.

Index: A range of scores, treated as interval or ratio level of measurement, measuring some phenomenon. In the example below, the higher one score the more politically conservative he or she is, 71.

Strongly Agree 20 Points

Agree 15

Unsure 10

Disagree 5

Strongly Disagree 0

Validity: The extent to which the concept one wishes to measure is actually being measured by a particular scale or index, 73.

Face Validity: The extent to which the measure is subjectively viewed by knowledgeable individuals as covering the concept.

Content Validity: The extent to which the measure covers all the generally accepted meanings of the concept.

Criterion Validity: The extent to which the measure is able to predict some criterion external to it, like a survey to assess future success in a position, 75.

Construct Validity: The ability of the scale to measure variables that are theoretically related to the variable that the scale purports to measure.

Reliability: The likelihood that the scale is actually measuring what it is supposed to measure, 76.

Split-half reliability: A measure of internal consistency that splits an overall scale into two scales, each containing half the original items.

Course Content

**Textbook Reading**

Sirkin, Chapters 1–3

Discussion Board

This session you are required to participate in two threads of the session-specific discussion board forum. Your

participation in both posting and responding to other students’ comments is graded. For this session’s discussion

topics, visit the discussion board in Blackboard.

Discussion Board Grading Scale:

**Critical Thinking --** We identify relevant information and cite our sources. We examine the arguments of others, identifying central issues and assumptions. We look at a question from various perspectives. We use correct logic in developing our argument.

**Contribution --** We do more than merely agree or disagree with the preceding discussion; we explain why we agree or disagree. We evaluate the previous discussion and draw inferences from it. And, most importantly, we add value to the discussion.

**Careful Writing --** We employ correct English spelling, grammar, syntax, and formatting, using APA style (the standard for the Predictive Analytics program) for writing and citations.  **Rushed or sloppy writing will not receive full credit.  It is your responsibility, not mine, to proofread your submissions.**

See below for information about writing assistance and writing references.

**Exposition --** This refers to the way we make our argument. We state the meaning or purpose of the post up front, perhaps telling the reader how many things will be discussed or how the material is organized. When making an argument, we make it clearly and concisely. Length is not automatically correlated with quality!

The discussions only on 'Prediction, Co-incidence & Causation' and 'Defining Variables' will be graded;

Research has shown that the likelihood of a child dying from SIDS is heightened for infants born to mothers who abused illegal drugs during pregnancy. Describe: 1) How and why using illegal drugs might cause a greater likelihood of SIDS, 2) How and why illegal drug use and SIDS might be correlated (that is, why greater levels of one occur with greater levels of the other) if, in fact, abusing drugs does not actually cause SIDS and 3) the relevance of this example to the field of predictive analytics.

1 The Gestation period is the closest physiological relationships a mother and baby will have in their relationship. During Gestation, the nutrient intake for a mother is shared with the baby through the placenta. As a result, when a mother ingests illegal drugs the baby also ingests the drugs. A baby’s development is much smaller on a scale compared to its mother and prone to nutrient risks. Illegal drugs are often filled with toxic nutrients for a baby and can often have a fatal reaction on the developing baby (John Hopkins Medicine).

Drugs are created with different substances, and some ingredients are more toxic than other ingredients depending on the drug. Mothers have different reactions to drugs and likewise, the baby will have reactions. Overall, given the negative effects drugs have on the Gestation period, professionals see a correlation between SIDS and drug use especially cocaine.

Reference

Johns Hopkins Medicine, based in Baltimore, Maryland. (n.d.). *Johns Hopkins Medicine, based in Baltimore, Maryland*. Retrieved June 20, 2012, from http://www.hopkinsmedicine.org

General Website Inquiry

2. How and why illegal drug use and SIDS might be correlated (that is, why greater levels of one occur with greater levels of the other) if, in fact, abusing drugs does not actually cause SIDS.

There are correlating factors that could link SIDS and drug abuse if it is found that ingesting drugs does not directly cause SIDS. According to John Hopkins Medicine, the healthier a mother is during her pregnancy the greater chances are her baby will be healthy. The inverse to this statement is the deterioration of a woman’s health increases the chances of a SIDS baby. While drug use might not directly cause SIDS, it could in fact lead to deteriorated health which is a factor in the causation of SIDS. Also, drug use is often associated with certain diseases because of ingesting techniques such as sharing needles. As a result, Hepatitis and immune deficiency diseases are associated with drug users. If a mother contracted one of the diseases and became pregnant it would greatly increase her chances for SIDS due to poor health and the suspected health of the developing baby. Drug abuse and poor health are often correlated. Likewise, poor health and SIDS are correlated. Drug abuse might not directly cause SIDS, but given the correlating relationship between drugs, poor health, and SIDS one could draw a connecting relationship.

Johns Hopkins Medicine, based in Baltimore, Maryland. (n.d.). *Johns Hopkins Medicine, based in Baltimore, Maryland*. Retrieved June 20, 2012, from http://www.hopkinsmedicine.org

the relevance of this example to the field of predictive analytics.

3. The example of mothers, drug use, and SIDS demonstrates a relationship. In predictive analytics, experiments are conducted to analyze and understand relationships. Our assigned reading taught how to form a hypothesis. The definition of a hypothesis according to R. Mark Sirkin is, “A statement that name the variables that appear to be related and indicates the nature of that relationship.(Page 10, 2006).” A hypothesis should have a dependent variable that is dependent or affected on an independent variable. In the case of the example for Session 1, the hypothesis would be worded, “there is a relationship between using drugs and SIDS such that when drugs are consumed by a mother there is a greater chance for SIDS.” From the hypothesis, an experiment can be conducted and analytics would help ascertain whether the statement is scientifically accurate. Analytics, relevant research and experiments help reveal pertinent information to issues like drug use and SIDS.

References

Sirkin, R. M. (2006). How We Reason. *Statistics for the social sciences* (3rd ed., p. 10). Thousand Oaks, Calif.: Sage Publications.

An important concept that people often lose sight of when interpreting data and what might be called coincidences is that “correlation is not causation” Discuss the meaning of this phrase. Why is it an important distinction to make? Describe some simple situations where two events coincide, or happen in succession, but neither is the direct cause of the other.

Why is precisely defining inputs and outputs important within statistics? For instance, how might a firm define “customer satisfaction” and why would the precision of such a definition be important?

3. Clearly articulated inputs and outputs are pivotal for the decision making process and communication plan for a firm. Precisely defined inputs and outputs create “a single version of the truth” for a firm (Davenport and Harris). Statistics seeks to understand and answer complex questions in a scientific manner in order to be profitable. In order to accurately gather information, the parameters of the experiment must be clearly defined and understood. Take for example, customer satisfaction as a desired inquiry. There are many different perspectives on the meaning of customer satisfaction. Management might perceive it to mean overall buying experience. Marketing might understand the term to mean likeliness to purchase in the next six months. If a costly experiment is conducted and the overall definition is not defined, major errors will result. Using statistics, one would have to clearly identify the meaning of customer satisfaction. Once established, an experiment would communicate input questions based on the definition to derive output driven results to help answer the defined inquiry. The result of such an experiment would prove beneficial to the overall success of a firm.

Davenport, T. H., & Harris, J. G. (2007). Analytics and Business Performance. *Competing on analytics the new science of winning* (p. 46). Boston, Mass: Harvard Business School Press.

The human condition makes it easy to assume correlation is causation. Perhaps early development of our species depended on the quick association for survival. In reality, the correlation of events does not prove causation. According to Webster’s Dictionary (2012) the definition of correlation is, “the state or relation of being correlated, specifically **:** a relation existing between phenomena or things or between mathematical or statistical variables which tend to vary, be associated, or occur together in a way not expected on the basis of chance alone.” The definition seems to make a link between correlation and causation, but the major flaw exacerbated by humans is the final part of the definition, “occur together in a way not expected on the basis of chance alone.” Linking events simply by correlation to causation leads to errors, which is why taking the phrase correlation is not causation to heart is very important.

Individuals may process information that lead them to believe events are correlating then jump to causation without doing due diligence. Take racial profiling for example; the extreme case of Henry Louis Gates Jr. (Boston Globe 2009) demonstrates the error of two events coinciding but having no causation relationship. In July of 2009, prominent African-American, Harvard professor Henry Louis Gates Jr. was arrested at his own home on the suspicion he was breaking into the home. Gates was in the process of dealing with a jammed door and police suspected he was breaking into his home. The police surmised the correlation of an African American male in that particular neighborhood struggling to open a door directly related to the causation of a crime being committed. Sadly, the police were massively mistaken, and it took the intervention of Barack Obama and barley pop, beer, to quell the situation.

References

Correlated - Definition and More from the Free Merriam-Webster Dictionary. (n.d.). *Dictionary and Thesaurus - Merriam-Webster Online*. Retrieved June 20, 2012, from http://www.merriam-webster.com/dictionary/correlated

Jan, T. (2009, July 20). Harvard professor Gates arrested at Cambridge home. *The Boston Globe*, p. 1. Retrieved June 20, 2012, from http://www.boston.com/news/local/breaking\_news/2009/07/harvard.html

Statistically the events leading up to Gates arrest appeared correlative to past experiences indicating a crime.

Download Info for SPSS

C:\Users\dprusins\AppData\Local\

**4-21152**

**Activation Code:** **SamplePower 3.0 (12 Month)**   
557b535a542247c41ad1

**Activation Code:** **Visualization Designer 1.0 (12 Month)**   
646fc8688820c30f27fb

**Activation Code:** **IBM SPSS Statistics Premium GradPack 20 (12-Mo Rental)**   
d67b54ccb56151eda9f5